



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

LLNL-JRNL-645399

# Complete genome sequence of *Micromonospora* L5, a plant-growth regulating actinomycete, originally isolated from *Casuarina equisetifolia* root nodules

M. Nolan, J. Han, N. Ivanova, A. Pati, T. Woyke, D. Bruce, L. Goodwin, A. Hirsch, J. Alvarado, O. Cherkov, P. De Hoff, C. Detter, N. Fujishige, S. Han, M. Land, M. Lum, C. Martinez, N. Milani-Nejad, S. Pitluck, M. Valdes

October 30, 2013

genomeA

## **Disclaimer**

---

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

# **Complete genome sequence of *Micromonospora* L5, a plant-growth regulating actinomycete, originally isolated from *Casuarina equisetifolia* root nodules**

<sup>a</sup>Ann M. Hirsch<sup>1,2</sup>, Johana Alvarado<sup>1</sup>, David Bruce<sup>4</sup>, Olga Chertkov<sup>3</sup>, Peter L. De Hoff<sup>2</sup>, Chris Detter<sup>4</sup>, Nancy A. Fujishige<sup>1</sup>, Lynne Goodwin<sup>4</sup>, James Han<sup>3</sup>, Shunsheng Han<sup>4</sup>, Natalia Ivanova<sup>3</sup>, Miriam L. Land<sup>5</sup>, Michelle R. Lum<sup>1</sup>, Claudia Martinez<sup>5</sup>, Nima Milani-Nejad<sup>1</sup>, Matt Nolan<sup>7</sup>, Amrita Pati<sup>3</sup>, Sam Pitluck<sup>3</sup>, Tanja Woyke<sup>3</sup>, and Maria Valdes<sup>6</sup>

<sup>1</sup>Department of Molecular, Cell and Developmental Biology

<sup>2</sup>Molecular Biology Institute, UCLA, Los Angeles, CA 90095-1606

<sup>3</sup>Lawrence Berkeley National Lab/Joint Genome Institute, Walnut Creek, CA

<sup>4</sup>Los Alamos National Laboratory, Los Alamos, NM

<sup>5</sup>Oak Ridge National Laboratory, Oak Ridge, TN

<sup>6</sup>Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, I.P.N., Plan de Ayala y Carpio, 11340 México, D.F.

<sup>7</sup>Lawrence Livermore National Lab/JGI, Walnut Creek, CA

*"To whom correspondence may be addressed. E-mail: [ahirsch@ucla.edu](mailto:ahirsch@ucla.edu)*

August 27, 2013

## **ACKNOWLEDGMENTS:**

The work conducted by the US Department of Energy (DOE) Joint Genome Institute is supported by the Office of Science of the DOE under Contract Number DE-AC02-05CH11231. Funding was also provided by UC-Mexus/CONACYT. Several UCLA undergraduate students are acknowledged for their help with the annotation: F. Candamo, I. Chou, H.S. Kim, S. Kim, W. Kim, M. Mojica, I. Ortiz, J.D. Pablo, R. Relatores, J. Terzian, J. Thai, C.M. Yasher, and J. Wong.

## **DISCLAIMER:**

[**LBNL**] This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific

## **Complete genome sequence of *Micromonospora* L5, a plant-growth regulating actinomycete, originally isolated from *Casuarina equisetifolia* root nodules**

commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California

**[LLNL]** This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

**Complete genome sequence of *Micromonospora* L5, a plant-growth regulating actinomycete, originally isolated from *Casuarina equisetifolia* root nodules.**

Ann M. Hirsch,<sup>a,b</sup> Johana Alvarado,<sup>a</sup> David Bruce,<sup>c</sup> Olga Chertkov,<sup>c</sup> Peter L. De Hoff,<sup>b</sup> Chris Detter,<sup>c</sup> Nancy A. Fujishige,<sup>a</sup> Lynne Goodwin,<sup>d</sup> James Han,<sup>d</sup> Shunsheng Han,<sup>c</sup> Natalia Ivanova,<sup>c</sup> Miriam L. Lande,<sup>e</sup> Michelle R. Lum,<sup>a</sup> Claudia Martinez,<sup>f</sup> Nima Milani-Nejad<sup>1</sup>, Matt Nolan<sup>5</sup>, Amrita Pati,<sup>c</sup> Sam Pitluck<sup>d</sup>, Tanja Woyke<sup>d</sup>, and Maria Valdes<sup>f</sup>

Department of Molecular, Cell and Developmental Biology<sup>1</sup> and Molecular Biology Institute, UCLA, Los Angeles, CA 90095-1606<sup>b</sup>; Joint Genome Institute, Walnut Creek, CA<sup>c</sup>; Los Alamos National Laboratory, Los Alamos, NM<sup>d</sup>; Oak Ridge National Laboratory, Oak Ridge, TN<sup>e</sup>; Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, I.P.N., Plan de Ayala y Carpio, 11340 México, D.F.<sup>f</sup>

Corresponding author: ahirsch@ucla.edu

Current addresses: JA, San Francisco Health Plan, 201 3rd Street, San Francisco, CA 94103; PLD, Synthetic Genomics, La Jolla, CA 92037; NAF, Dept. of Biology, Loyola Marymount University, Los Angeles, CA 90045; MRL, Dept. of Biology, Loyola Marymount University, Los Angeles, CA 90045; NM-N, Medical Scientist Training Program, Biomedical Sciences Graduate Program, Department of Physiology and Cell Biology, D. Davis Heart and Lung Research Institute, College of Medicine, The Ohio State University, Columbus, Ohio, 43210

Abstract



*Micromonospora* species live in diverse environments including within legume and actinorhizal root nodules (1-4). Very few *Micromonospora* genomes have been fully sequenced even though many species enhance plant growth (5). Here, we contribute to this effort with the complete genome sequence of *Micromonospora* strain L5, which was isolated from nodules of *Casuarina equisetifolia* trees growing in Mexico (1, 2).

The *Micromonospora* L5 genome was sequenced at the Joint Genome Institute (JGI) using a combination of Illumina (6) and 454 technologies (7). An Illumina GAii shotgun library with reads of 868 Mb, a 454 Titanium draft library with average read length of 510-525 bp bases, and paired end 454 libraries with average insert sizes of 10 Kb and 14 Kb were generated for this genome. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. Illumina sequencing data was assembled with VELVET (8), and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. Draft assemblies were based on 793.7 MB 454 draft data, and all of the 454 paired end data. Newbler parameters are - consed -a 50 -l 350 -g -m -ml 20. The initial Newbler assembly contained 181 contigs in 8 scaffolds.

We converted the initial 454 assembly into a phrap assembly by making fake reads from the consensus, collecting the read pairs in the 454 paired end library. The Phred/Phrap/Consed software package ([www.phrap.com](http://www.phrap.com)) was used for sequence assembly and quality assessment (9-11) in the following finishing process. Illumina data were used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI (Alla Lapidus, unpublished). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies

were corrected with gapResolution (Cliff Han, unpublished), Dupfinisher (12), or sequencing cloned bridging PCR fragments with sub-cloning. Gaps between contigs were closed by editing in Consed, by PCR, and by Bubble PCR primer walks. A total of 732 additional reactions and 4 shatter libraries were necessary to close gaps and to raise the quality of the finished sequence.

The genome has a size of 6,907,073 bp, 6332 predicted ORFs, and a GC content of 72.86%. It most likely has a circular chromosome based on its close relationship to *M. aurantiaca* (13). We found 4248 known protein and 1984 hypothetical ORFs, 52 tRNAs, and 2 rRNAs. The largest number of orthologs was shared with *M. aurantiaca* ATTC 27029 (NCBI Bioproject [PRJNA42501](#)). Although strain L5 was reported to fix nitrogen via numerous physiological tests (14), we could not find bona fide *nif* gene sequences in the L5 genome. Genes were found for the synthesis of a broad range of cell-wall degrading enzymes, a second independent system, the Tat (twin arginine translocation) export pathway, and types II and VII secretion systems.

Nucleotide sequence accession numbers. The complete sequence of *Micromonospora* L5 has been deposited at NCBI Genbank as NCBI Bioproject PRJNA38291.

## 67    **Acknowledgements**

68    The work conducted by the U.S. Department of Energy Joint Genome Institute is supported  
69    by the Office of Science of the U.S. Department of Energy Under Contract No. DE-AC02-  
70    05CH11231.

71    Funding was also provided by UC-Mexus/CONACYT. Several UCLA undergraduate students  
72    are acknowledged for their help with the annotation: F. Candamo, I. Chou, H.S. Kim, S. Kim,  
73    W. Kim, M. Mojica, I. Ortiz, J.D. Pablo, R. Relatores, J. Terzian, J. Thai, C.M. Yasher, J. Wong,  
74    and .

75

Prepared by LLNL under Contract DE-AC52-07NA27344.

## References

1. **Guillén G, Valdés M, Liao J, Hirsch AM.** 1993. Identificación de actinobacterias aisladas de nódulos de *Casuarina*, por técnicas tradicionales y moleculares. *Revista Lat. Amer. Microbiol.* **35**:195-200.
2. **Niner BM, Brandt JP, Villegas M, Marshall CR, Hirsch AM, Valdés, M.** 1996. Analysis of partial sequences of genes coding for 16S rRNA of actinomycetes isolated from *Casuarina equisetifolia* nodules in Mexico. *Appl. Environ. Microbiol.* **62**:3034-3036.
3. **Trujillo ME, Kroppenstedt RM, Schumann P, Carro L, Martínez-Molina E.** 2006. *Micromonospora coriariae* sp. nov., isolated from root nodules of *Coriaria myrtifolia*. *Inter. J. Syst. Evol Microbiol.* **56**:2381-2385.
4. **Trujillo ME, Kroppenstedt, R.M., Fernández-Molineo C, Schumann P, Martínez-Molina E.** 2007. *Micromonospora lupini* sp. nov. and *Micromonospora saelicesensis* sp. nov., isolated from root nodules of *Lupinus angustifolius*. *Inter. J. Syst. Evol Microbiol.* **57**: 2799-2804.
5. **Hirsch AM, Valdés M.** 2010. *Micromonospora*: a versatile microbe not only for biomedicine, but also for biocontrol and the production of biofuels. *Soil Biol. Biochem.* **42**:536-542.
6. **Bennett S.** Solexa Ltd. Pharmacogenomics. 2004. 5:433-438.
7. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H,**

- Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437:376-80.
8. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 18:821-829.
9. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 8:186-194.
10. **Ewing B, Hillier L, Wendl MC, Green P.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* 8:175-185.
11. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. *Genome Res*. 8:195-202.
12. **Han C, Chain P.** 2006. Finishing repeat regions automatically with Dupfinisher. In *Proceeding of the 2006 international conference on bioinformatics & computational biology*. Edited by Hamid R. Arabnia & Homayoun Valafar, CSREA Press. June 26-29, 2006: 141-146.
13. **Kirby R.** 2011. Chromosome diversity and similarity within the *Actinomycetales*. *FEMS Microbiology Lett*. 319:1-10.
14. **Valdés M, Pérez N-O, Estrada-de los Santos P, Caballero-Mellado J, Peña-Cabriaes JJ, Normand P, Hirsch AM.** 2005. Non-*Frankia* actinomycetes isolated from surface-

119 sterilized roots of *Casuarina equisetifolia* fix nitrogen. Appl. Environ. Microbiol. **71**:  
120 460-466.

121

122